

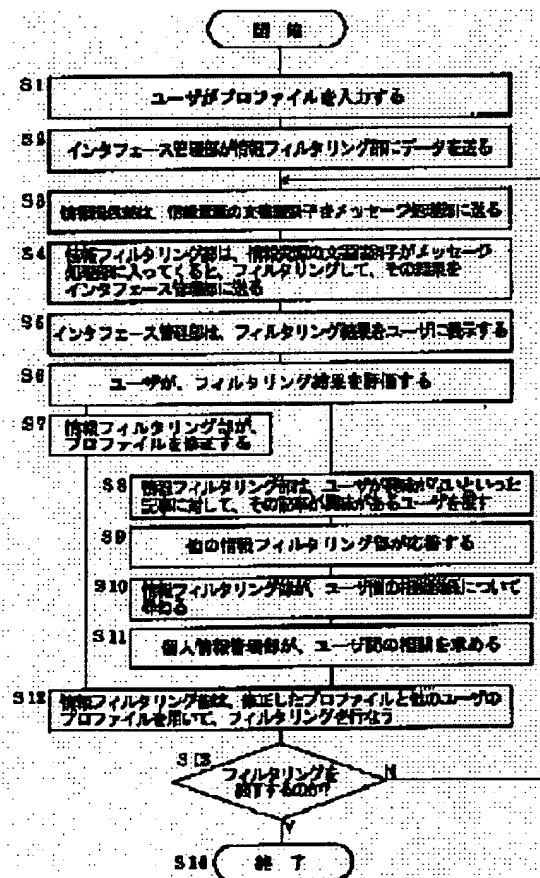
INFORMATION FILTERING METHOD AND DEVICE

Patent number: JP11282874
Publication date: 1999-10-15
Inventor: SUGAI TAKESHI; HANIYUDA HIROMI; MORITA KOHAKU
Applicant: OKI ELECTRIC IND CO LTD
Classification:
- international: G06F17/30
- european:
Application number: JP19980100168 19980327
Priority number(s): JP19980100168 19980327

Report a data error here

Abstract of JP11282874

PROBLEM TO BE SOLVED: To obtain a highly accurate filtered result even when a profile is a polysemous work. **SOLUTION:** When a profile is inputted by as user (a step S1), filtering based on the profile is performed and its result is displayed to the user by an information filtering part (a step S5). The filtered result is evaluated by the user (a step S6), and when information resources in which no interest is shown by the user exist, another user having interest in the information resources is searched by the information filtering part (a step S8). Next, correlation between such other user and the user as a requester for filtering is searched. When no correlation exists between both users, the filtered result by the other user's profile is excluded from the result outputted in the step 5 and is outputted as a new filtered result (a step S12).



Data supplied from the esp@cenet database - Worldwide

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-282874

(43)公開日 平成11年(1999)10月15日

(51)Int.Cl.⁶

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/403

3 4 0 A

15/40

3 1 0 F

3 1 0 C

審査請求 未請求 請求項の数 5 F D (全 10 頁)

(21)出願番号 特願平10-100168

(22)出願日 平成10年(1998)3月27日

(71)出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72)発明者 菅井 猛

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72)発明者 羽生田 博美

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72)発明者 森田 幸伯

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

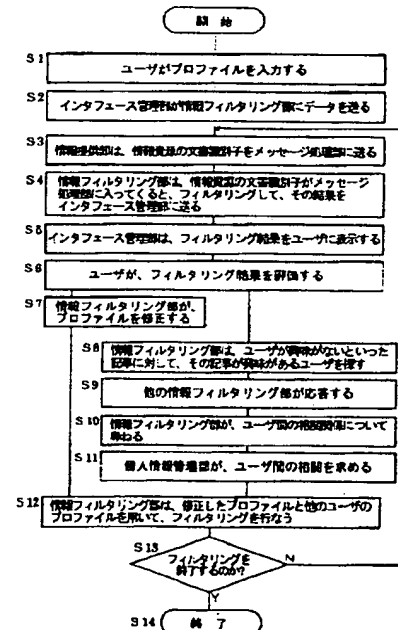
(74)代理人 弁理士 佐藤 幸男 (外1名)

(54)【発明の名称】 情報フィルタリング方法および装置

(57)【要約】

【課題】 プロファイルが多義語であっても、精度の高いフィルタリング結果を得る。

【解決手段】 ユーザがプロファイルを入力すると(ステップS1)、情報フィルタリング部は、このプロファイルに基づくフィルタリングを行い、結果をユーザに表示する(ステップS5)。ユーザがフィルタリング結果を評価し(ステップS6)、ユーザが興味がないといった情報資源があった場合、情報フィルタリング部は、この情報資源に対して興味のある他のユーザを探す(ステップS8)。次に、このような他のユーザとフィルタリングの要求者であるユーザとの相関関係を求める。双方のユーザに相関関係がなかった場合、他のユーザのプロファイルによるフィルタリング結果をステップS5で出力した結果から除外してあたらなフィルタリング結果として出力する(ステップS12)。



本発明方法の具体例1のフローチャート

【特許請求の範囲】

【請求項1】 ユーザのプロファイルに適合した情報資源を収集すると共に、

同じプロファイルを異なる意味で用いている他のユーザを抽出し、当該他のユーザと前記ユーザとの相関関係を求め、相関関係がない場合は、前記他のユーザが興味がある情報資源を、前記収集した情報資源から除外して前記ユーザへのフィルタリング結果とする情報フィルタリング方法。

【請求項2】 ユーザのプロファイルに適合した情報資源を収集してユーザに出力し、

その結果、前記出力した情報資源のうち、前記ユーザが興味が無いといった情報資源に対して、興味があり、かつ、前記ユーザと同じプロファイルを用いている他のユーザを抽出し、

次いで、前記ユーザと他のユーザとの相関関係を求め、相関関係がない場合は、前記他のユーザが興味のある情報資源を、前記収集した情報資源から除外して前記ユーザへの新たなフィルタリング結果とする情報フィルタリング方法。

【請求項3】 請求項1または2に記載の情報フィルタリング方法において、

ユーザと他のユーザとの相関関係は、ソーシャルフィルタリングを行うことで判定することを特徴とする情報フィルタリング方法。

【請求項4】 複数のユーザの属性情報を格納するユーザ情報格納部と、

前記ユーザ情報格納部に格納された複数のユーザの属性に対する相関関係を示す情報を管理する個人情報管理部と、

特定のユーザが入力したプロファイルを含む情報資源をフィルタリング結果として出力すると共に、当該フィルタリング結果に対して、前記特定のユーザが興味がないと応答した情報資源があった場合、当該情報資源に興味があり、かつ、前記特定のユーザと同じプロファイルを用いている他のユーザを抽出し、当該他のユーザが前記特定のユーザと相関関係があるかを前記個人情報管理部に問い合わせ、当該個人情報管理部より相関関係がないという応答があった場合は、前記特定のユーザのプロファイルを含む情報資源から、前記他のユーザが興味があるとした情報資源を除外して前記特定のユーザへの新たなフィルタリング結果とする情報フィルタリング部とを備えたことを特徴とする情報フィルタリング装置。

【請求項5】 任意のユーザが特定の情報資源に対して評価した情報を格納するユーザ評価情報格納部と、情報フィルタリングの要求者であるユーザと、他の複数のユーザが、同一の情報資源に対して評価を行い、この評価値に基づき、前記他の複数のユーザのうち、前記要求者であるユーザと類似度の最も低いユーザを抽出するソーシャルフィルタリング部と、

特定のユーザが入力したプロファイルを含む情報資源をフィルタリング結果として出力すると共に、当該フィルタリング結果に対して、前記特定のユーザが興味がないと応答した情報資源があった場合、当該情報資源に興味があり、かつ、前記特定のユーザと同じプロファイルを用いている他のユーザを抽出し、当該他のユーザのうち、前記特定のユーザと最も類似度が低いユーザが誰であるかを前記ソーシャルフィルタリング部に問い合わせ、当該ソーシャルフィルタリング部が応答した最も類似度の低いユーザが興味があるとした情報資源を、前記特定のユーザのプロファイルを含む情報資源から除外して新たなフィルタリング結果とする情報フィルタリング部とを備えたことを特徴とする情報フィルタリング装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報資源に対してフィルタリングを行う情報フィルタリング方法および装置に関し、特に、ユーザの嗜好に合わせてユーザのプロファイルを多段に組み合わせる情報フィルタリング方法および装置に関する。

【0002】

【従来の技術】従来の情報フィルタリングの技術を示すものとして、以下の文献があった。

【1】Nicholas J. Belkin, W. Bruce Croft, Information Filtering and Information retrieval: Two Sides of the Same Coin?, Communication of the ACM, 35(12), pp29-38, 1992

【0003】【2】Upendra Shardanand, Pattie Maes, Social Information Filtering: Algorithms for Automating "Word of Mouth", CHI '95 Conference Proceedings, 1995, pp210-217

【0004】【3】Chris Beckley, Gerard Salton, James Allan, The Effect of Adding Relevance Information in a Relevance Feedback Environment, SIGIR '94 Proceedings, 1994, pp292-300

【0005】コンピュータネットワークの整備によって、電子メールやネットニュース等、日々刻々と変わる情報に対して、ユーザが欲しい情報だけを得るという情報フィルタリングの技術の開発が行われている。情報フィルタリングでは、ユーザの日々少しずつ変わる興味に追従するために、関連フィードバック (Relevance Feedback) や学習アルゴリズムによって、ユーザのプロファイルを修正してフィルタリングを行うシステムが一般的である。

【0006】ここで、プロファイルとは情報検索における検索文、あるいは、質問文としてはほぼ同様の意味であるが、ここでは、ある時間の長さに対して、ユーザの検索の興味を表現した検索文である。こうしたフィルタリングは、テキストの内容とユーザが欲する情報を表現し

たプロファイルのパターンマッチによって行われ、これは、内容に基づくフィルタリングと呼ばれている（これについては、例えば上記文献[1]に示す）。

【0007】情報フィルタリングの手順は、一般に以下のような方法で行われる。

1. ユーザは、初期プロファイルを入力する。
2. 情報フィルタリング装置は、ユーザのプロファイルと配送されてきたテキストを比べて、フィルタリングを行い、フィルタリング結果をユーザに表示する。
3. ユーザは、フィルタリング結果に対する評価を入力する。
4. 2～4を繰り返す。

【0008】

【発明が解決しようとする課題】しかしながら、上記従来の情報フィルタリング装置では、次のような問題点があった。先ず、情報フィルタリングの手順においては、ユーザはフィルタリングされた記事に対して、興味があるか、あるいは、興味がないかを入力するが、フィルタリング手順の上記2～4の繰り返し回数が少ない段階では、プロファイルが十分な修正が行われておらず、フィルタリングの精度があまりよくならないという問題があった。

【0009】次の問題点として、プロファイルは、情報資源をフィルタリングするには十分な情報を一般的には含んでいない。従って、プロファイル中の特定の情報（ある基底語、あるいはある特徴ベクトル）は、データの分野毎に重要度が異なる。従って、同一のプロファイルで多分野のデータをフィルタリングしようとする、ある特定の分野では重要度が高いが別の分野では重要度が低いようなプロファイルの基底語が存在し、全体としてフィルタリングの精度が低下する。尚、基底語とは、発明者が先に出願した特願平9-157909号明細書に記載したように、ベクトル空間モデルにおけるベクトルの軸を構成するものである。また、ベクトル空間モデルについては、実施態様の項で詳述する。

【0010】例えば、ベクトル空間モデルでは、多義語は同じ特徴ベクトルとして表現されてしまうので、検索精度が悪くなる一つの要因であった。その一例としては、コンピュータのデータベースエンジニアが「モデル」というプロファイルを指定し、フィルタリング結果に、ファッションに関する「モデル」のテキストが得られたとする。ここで、「モデル」という語は、文脈によって異なった意味を持つので多義語である。これを関連フィードバックによりプロファイルを洗練しようとする、と、「モデル」という基底語の重要度が低くなってしまい、次のフィルタリングの過程では、データベースの「モデル」という基底語を含むテキストをフィルタリングすることができなくなってしまう。

【0011】更に、フィルタリングの場合、分野別に区分けされていないデータベースを対象とすることが多

く、このようなデータベースに対して多義語によるフィルタリングの精度を向上させるのは困難であった。

【0012】このような点から、プロファイルが多義語であっても、精度の高いフィルタリングが行える情報フィルタリング方法および装置の実現が望まれていた。

【0013】

【課題を解決するための手段】本発明は、前述の課題を解決するため次の構成を採用する。

〈請求項1の構成〉ユーザのプロファイルに適合した情報資源を収集すると共に、同じプロファイルを異なる意味で用いている他のユーザを抽出し、他のユーザとユーザとの相関関係を求め、相関関係がない場合は、他のユーザが興味がある情報資源を、収集した情報資源から除外してユーザへのフィルタリング結果とする情報フィルタリング方法である。

【0014】〈請求項1の説明〉同じプロファイルを異なる意味で用いているとは、例えば、コンピュータのデータベースに関する「モデル」に対して、ファッションに関する「モデル」といったように、多義語のプロファイルにおいて異なる意味で用いる場合を指している。ユーザ間の相関関係とは、例えば、ユーザの職業に基づき相関関係があるか否かを判定するが、予め決められた条件であれば、職業以外のものを基準にして相関関係を判定するようにしてもよい。

【0015】即ち、請求項1の発明は、あるプロファイルのフィルタリング結果から、自分と同じプロファイルを違う意味で用いており、かつ、自分とは相関関係のないユーザのフィルタリング結果を除外して、新たなフィルタリング結果としたものである。

【0016】請求項1の発明がこのように構成されていることにより、ユーザに関係のない情報資源を削除することができ、その結果、フィルタリングの精度を向上させることができる。

【0017】〈請求項2の構成〉ユーザのプロファイルに適合した情報資源を収集してユーザに出力し、その結果、出力した情報資源のうち、ユーザが興味が無いといった情報資源に対して、興味があり、かつ、ユーザと同じプロファイルを用いている他のユーザを抽出し、次いで、ユーザと他のユーザとの相関関係を求め、相関関係がない場合は、他のユーザが興味のある情報資源を、収集した情報資源から除外してユーザへの新たなフィルタリング結果とする情報フィルタリング方法である。

【0018】〈請求項2の説明〉請求項2の発明は、請求項1の発明における、同じプロファイルを異なる意味で用いている他のユーザの抽出方法として、先ず、初期プロファイルによるフィルタリング結果を出力し、次に、この結果に対してユーザが興味がないといった情報資源に興味があり、かつ、同じプロファイルを用いているユーザをこのような他のユーザとして抽出するようにしたものである。

【0019】このような構成により、ユーザがプロフィールを修正する等の操作を繰り返さなくても、精度の高い検索結果を得ることができる。

【0020】〈請求項3の構成〉請求項1または2に記載の情報フィルタリング方法において、ユーザと他のユーザとの相関関係は、ソーシャルフィルタリングを行うことで判定することを特徴とする情報フィルタリング方法である。

【0021】〈請求項3の説明〉請求項3の発明は、請求項1または2の発明において、ソーシャルフィルタリングの機能を用いて、ユーザ間の相関関係を判定するようにしたものである。これにより、請求項1または2の発明と同様の効果を得ることができる。

【0022】〈請求項4の構成〉複数のユーザの属性情報を格納するユーザ情報格納部と、ユーザ情報格納部に格納された複数のユーザの属性に対する相関関係を示す情報を管理する個人情報管理部と、特定のユーザが入力したプロフィールを含む情報資源をフィルタリング結果として出力すると共に、フィルタリング結果に対して、特定のユーザが興味がないと応答した情報資源があった場合、情報資源に興味があり、かつ、特定のユーザと同じプロフィールを用いている他のユーザを抽出し、他のユーザが特定のユーザと相関関係があるかを個人情報管理部に問い合わせ、個人情報管理部より相関関係がないという応答があった場合は、特定のユーザのプロフィールを含む情報資源から、他のユーザが興味があるとした情報資源を除外して特定のユーザへの新たなフィルタリング結果とする情報フィルタリング部とを備えたことを特徴とする情報フィルタリング装置である。

【0023】〈請求項4の説明〉ユーザの属性情報は、例えば、コンピュータエンジニアやファッション業界で働く人といった属性であるが、ユーザを予め決められた条件で分類するものであればどのようなものであってもよい。また、ユーザの属性に対する相関関係を示す情報とは、例えば、「ファッション関係で働く人」と「コンピュータ業界で働く人」との関係は、「あまり強い相関がない」といった情報であるが、相関関係を示す情報であればどのような情報であってもよい。

【0024】情報フィルタリング部は、まず、特定のユーザが入力したプロフィールに基づくフィルタリング結果を出力する。そして、特定のユーザはこの結果に対して、興味の有無を入力する。これにより、情報フィルタリング部は、特定のユーザが興味がないと入力した情報資源に対して、興味があり、かつ、同じプロフィールを用いている他のユーザを抽出する。次に、情報フィルタリング部は、他のユーザと特定のユーザとの相関関係を個人情報管理部に問い合わせる。

【0025】これにより、個人情報管理部は、ユーザ情報格納部のユーザ属性情報を参照し、双方のユーザに相関関係があるかを調べ、その結果を情報フィルタリング

部に通知する。「相関関係がない」の結果であった場合、情報フィルタリング部は、特定のユーザのプロフィールを含む情報資源から、他のユーザが興味があるとした情報資源を除外して新たなフィルタリング結果として出力する。

【0026】請求項4の発明は、このように構成されていることにより、ユーザがプロフィールを修正する等の操作を繰り返さなくても、精度の高い検索結果を得ることができる。

【0027】〈請求項5の構成〉任意のユーザが特定の情報資源に対して評価した情報を格納するユーザ評価情報格納部と、情報フィルタリングの要求者であるユーザと、他の複数のユーザが、同一の情報資源に対して評価を行い、この評価値に基づき、他の複数のユーザのうち、要求者であるユーザと類似度の最も低いユーザを抽出するソーシャルフィルタリング部と、特定のユーザが入力したプロフィールを含む情報資源をフィルタリング結果として出力すると共に、フィルタリング結果に対して、特定のユーザが興味がないと応答した情報資源があった場合、情報資源に興味があり、かつ、特定のユーザと同じプロフィールを用いている他のユーザを抽出し、他のユーザのうち、特定のユーザと最も類似度が低いユーザが誰であるかをソーシャルフィルタリング部に問い合わせ、ソーシャルフィルタリング部が応答した最も類似度の低いユーザが興味があるとした情報資源を、特定のユーザのプロフィールを含む情報資源から除外して新たなフィルタリング結果とする情報フィルタリング部とを備えたことを特徴とする情報フィルタリング装置である。

【0028】〈請求項5の説明〉請求項5の発明は、特定のユーザが興味がないと応答した情報資源に対して興味があり、かつ、特定のユーザと同じプロフィールを用いている他のユーザのうち、特定のユーザと最も類似度が低いユーザが誰であるかを、ソーシャルフィルタリングの機能を用いて判定するようにしたものである。このような構成により、請求項4の発明と同様、ユーザがプロフィールを修正する等の操作を繰り返さなくても、精度の高い検索結果を得ることができる効果がある。

【0029】

【発明の実施の形態】以下、本発明の実施の形態を図面を用いて詳細に説明する。

《具体例1》

〈構成〉図1は本発明の具体例1による情報フィルタリング方法を示すフローチャートであるが、この説明に先立ち、情報フィルタリング方法を実施するための情報フィルタリング装置の説明を行う。

【0030】図2は本発明の情報フィルタリング装置の具体例1を示す構成図である。図の装置は、ユーザ情報格納部1、個人情報管理部2、インタフェース管理部3、情報フィルタリング部4、メッセージ処理部5、情

報提供部6からなり、これら構成はコンピュータ上に実現されている。

【0031】ユーザ情報格納部1は、コンピュータのハードディスク装置等の記憶装置に設けられ、複数のユーザの属性情報を格納する格納部である。ここで、ユーザの属性とは、例えばユーザの職業といったものである。その一例としては、あるユーザは「ファッション関係の会社に勤める人」であり、他のユーザは「コンピュータエンジニア」といった情報である。

【0032】個人情報管理部2は、ユーザ情報格納部1に格納された複数のユーザの属性間の相関関係を示す情報を管理する機能を有している。例えば、「ファッション業界で働く人」と「コンピュータ業界で働く人」との関係は「あまり相関がない」といった情報を有している。

【0033】インタフェース管理部3(3a、3b、3c)は、複数のユーザ20a、20b、20c毎に設けられ、各ユーザの入力を受け付け、そのユーザが入力した情報を情報フィルタリング部4に送り、また、情報フィルタリング部4から送られてきたフィルタリング結果を対応したユーザに表示する機能を有している。

【0034】情報フィルタリング部4(4a、4b、4c)は、インタフェース管理部3に対応して各ユーザ毎に設けられ、対応したユーザのプロファイルと情報源10とを比較して、パターンマッチに基づくフィルタリングを行う機能と、ユーザのフィードバックによりプロファイルを修正する機能を持つ。また、ユーザが入力したプロファイル文を基底語に変換し、特徴ベクトルを生成する機能を有している。

【0035】即ち、情報フィルタリング部4は、特定のユーザ(自情報フィルタリング部4に対応したユーザ)が入力したプロファイルを管理し、このプロファイルに基づき、情報源10からフィルタリングを行うと共に、このフィルタリング結果を出力したことによる特定のユーザからの興味の有無によりプロファイルを修正し、かつ、特定のユーザが興味がないと指示した情報資源に対して、興味があるユーザを探し、このユーザが特定のユーザと相関関係があるかを個人情報管理部2に問い合わせ、個人情報管理部2より相関関係がないという応答があった場合は、特定のユーザのプロファイルによるフィルタリング結果から、他のユーザのプロファイルでフィルタリングした結果を削除して特定のユーザへのフィルタリング結果とする機能を有している。

【0036】尚、図面上では、ユーザ20、インタフェース管理部3、情報フィルタリング部4は、それぞれ三つしか示していないが、実際には4人以上のユーザが存在し、かつ、これらのユーザに対応してインタフェース管理部3および情報フィルタリング部4が設けられているとする。

【0037】メッセージ処理部5は、情報フィルタリン

グ部4、情報提供部6、個人情報管理部2の通信の場を提供する機能を有している。また、情報提供部6は、情報源10中の情報資源の識別子(Identifier(id))をメッセージ処理部5に送る機能を有している。ここで、情報資源とは、情報源10に含まれる情報の一単位を示し、識別子は、各情報資源を識別するための情報である。また、本具体例では、この識別子を文書識別子と呼ぶ。

【0038】情報源10は、上述したように複数の情報資源からなり、文書やイメージのデータを含んでいる。そして、フィルタリング装置は、コンピュータネットワークを通じて情報源10に接続されている。

【0039】上記のインタフェース管理部3～情報提供部6は、それぞれの機能をコンピュータ上で実現するためのプログラムと、これらのプログラムを実行するプロセッサやメモリ等により各機能が構成されている。

【0040】〈動作〉具体例1では、ユーザ(これをユーザ20aとする)が、コンピュータのデータベースに関する本の中で、「モデル」に関する雑誌の記事をフィルタリングしたいとする。ユーザ20aがフィルタリング装置にプロファイルを登録しておけば、登録後、情報フィルタリング装置は、「モデル」に関する雑誌の記事が、ある情報資源の中に入ってくる毎に、その雑誌の記事をユーザに配送する。

【0041】以下、具体例1の動作を図1に沿って説明する。

【0042】〔ステップS1〕ユーザの起動コマンドにより、情報フィルタリング装置が起動する。この時、情報フィルタリング装置は、ユーザ20のデータの入力待ちとなる。今、ユーザ20aがプロファイルとして、「モデル」という語をインタフェース管理部3aを介して入力したとする。また、ユーザ20aの個人情報としては、「コンピュータエンジニア」であることを入力する。

【0043】〔ステップS2〕インタフェース管理部3aは、上記ステップS1で入力されたプロファイルを、情報フィルタリング部4aに送る。情報フィルタリング部4aは、ユーザ20aのプロファイルから特徴ベクトルを作成する。

【0044】ここで、情報フィルタリング部4aがフィルタリングを行うベクトル空間モデル、関連フィードバック(Relevance Feedback)について説明する。尚、ベクトル空間モデル、関連フィードバックは、情報フィルタリングを行うための一つの手法として用いたものであり、これ以外にも、他の方法、例えば情報検索における確率モデル、機械学習を用いてプロファイルを洗練するといったことなどでも実現が可能である。

【0045】ベクトル空間モデルはテキストの中の単語をベクトルとしてとらえる検索手法である(例えば、これについては文献[3]に示す)。テキストを単語に分

割して、その単語に重要度を割り当て、特徴ベクトルとする。ここで、一般に、分散ネットワーク上では、情報資源とは、画像データ、動画データ、圧縮ファイルなどを含むが、本具体例では、情報資源をテキストに限定する。また、質問文には、自然言語文を用い、テキスト（あるいは文書）の単語と同様、特徴ベクトルに変換する。検索結果は、テキストと質問文の類似度をランキングしたものである。ここで、以後、単語を基底語と呼ぶ【0046】あるテキストのベクトルを次式で表現する。図3は、ベクトル空間モデルの演算式の説明図である。図3中、(1)はあるテキストのベクトルDWを示す式である。(1)において、 dw_1, dw_2, \dots, dw_n は、そのテキストの基底語の重要度である。

【0047】同様に、質問文のベクトルQを(2)式で表現する。ここで、 q_1, q_2, \dots, q_n は、質問文の基底語の重要度である。

【0048】また、基底語の重要度(文書D、の語 T_i のベクトルの重み)は、(3)式で与えられる。ここで、各記号の意味は以下の通りである。

- ・ W_{ij} : 文書D、の基底語 T_i のベクトルの重み
- ・ $t_{f_{ij}}$: 文書D、の基底語 T_i の出現数
- ・N: 収集された文書の総数
- ・ n_i : 収集された文書の中で、基底語 T_i が含まれている文書の数

【0049】また、類似度Simは、(4)に示す式となる。ここで、(5)中の θ のようなスレッシュホールドが存在する。検索結果は、(3)式を満たしたものの中で類似度が大きい順序に表示される。

【0050】検索された文書をユーザが評価して、ユーザが関連あるといった文書のベクトルを質問文にフィードバックをかけて、検索式を洗練する方法を関連フィードバックという(この関連フィードバックについては、例えば文献[3]に示されている)。

【0051】関連フィードバックは様々な方法が提案されているが、一般に、例えば、文献[3]に示すような

(6)式が用いられる。ここで、 rel_docs とは、検索された文書の中で、ユーザが興味のある文書の特徴ベクトルである。また、 $nonrel_docs$ とは、検索された文書の中で、ユーザが興味がない文書の特徴ベクトルである。一般には、 α, β, γ に、値8、16、4をそれぞれ用いる。

【0052】【ステップS3】情報提供部6は、情報源10から送られてきた情報資源の文書識別子をメッセージ処理部5に送る。

【0053】【ステップS4】情報フィルタリング部4aは、情報資源の文書識別子がメッセージ処理部5に入ってくると、フィルタリングしてその結果をインタフェース管理部3aに送る。ここで、フィルタリングは、上記ベクトル空間モデルの式(1)～(5)で行われる。

【0054】【ステップS5】インタフェース管理部3

aは、フィルタリング結果をユーザに表示する。

【0055】図4は、情報フィルタリング部4aの動作の説明図である。この図の状態は、上記のステップS1～S5までの処理の流れを示すものである。即ち、情報源10の中から、情報資源の識別子である文書識別子が情報提供部6を介して情報フィルタリング部4aに入力される。情報フィルタリング部4aは、このような文書識別子に対して、ユーザ20aのプロファイルを参照してフィルタリングを行い、フィルタリング結果を出力する。

【0056】【ステップS6】ユーザ20aがフィルタリング結果を評価する。ここでは、ユーザ20aは、各情報資源に対して、その情報が興味があるかないかを入力する。

【0057】【ステップS7】情報フィルタリング部4aが、ステップS6におけるユーザ20aの評価情報を元に、プロファイルを修正する。ここでは、プロファイルの修正を、関連フィードバックの式(6)で行う。

【0058】【ステップS8】情報フィルタリング部4aは、ユーザ20aが興味がないと入力した記事に関して、その記事に興味があるユーザがいるかどうかをメッセージ処理部5に問い合わせる。

【0059】【ステップS9】他の情報フィルタリング部4bが、ステップS8のメッセージに回答したとする。ここでは、情報フィルタリング部4bが回答したとして、そのメッセージをメッセージ処理部5に送る。

【0060】【ステップS10】情報フィルタリング部4aは、ステップS9で回答した情報フィルタリング部4bについて、ユーザ間の相関を個人情報管理部2に尋ねる。

【0061】【ステップS11】個人情報管理部2が、ユーザ間の相関を答える。ここでは、個人情報管理部2はステップS10のメッセージに対して、ユーザ20aとユーザ20bとは「強い相関がない」ということをメッセージ処理部5に送る。

【0062】ここで、ユーザ間の相関を求める具体的な方法を説明する。図5は、ユーザの分類を示す説明図である。図示のように、ユーザは「コンピュータ」、「ファッション」、「経済」等に分類され、「コンピュータ」に属するユーザは、「ヒューマンインタフェース」、「データベース」、「コンピュータネットワーク」などのユーザに分類される。例えば、このユーザ20aは、「コンピュータ」という分類の中の「データベース」という分類の中に入る。

【0063】個人情報管理部2では、図5に示すような関係において、ユーザの分類が一番上のレベルで異なっていたら、そのユーザ間は、相関がないとする。上記例の場合、ユーザ20aは「コンピュータ」の分類に属しており、一方、ユーザ20bは「ファッション」の分類に属しているので、ユーザ20aとユーザ20bとは相

関がないといえる。

【0064】〔ステップS12〕情報フィルタリング部4aは、修正したプロフィールと他のユーザ20bのプロフィールを用いて、フィルタリングを行う。ここでは、情報フィルタリング部4aは、情報フィルタリング部4bのプロフィールによってフィルタリングされる情報資源を除いてフィルタリングを行う。このとき、ユーザ20a、インタフェース管理部3a、情報フィルタリング部4a、4b、情報提供部6、情報源10の関係は以下ようになる。

【0065】図6は、フィルタリングの順番を示す説明図である。図示のように、情報提供部6より提供された情報源10の文書識別子（情報資源）に対して、先ず、情報フィルタリング部4bのプロフィールによるフィルタリング（除外のフィルタリング）を行い、次いで、情報フィルタリング部4aのプロフィールによるフィルタリングを行って、インタフェース管理部3aを介してユーザ20aに結果が出力される。

【0066】〔ステップS13〕ユーザ20aがフィルタリングを終了するかを入力する。

【0067】〔ステップS14〕ステップS13において、ユーザ20aがフィルタリングを終了することを入力したならば、フィルタリング装置はフィルタリングを終了する。

【0068】〈効果〉以上のように、具体例1によれば、ユーザのプロフィールの中の基底語（あるいはキーワード）でその基底語を違う意味のプロフィールとして用いている、他のユーザのプロフィールを探して、そのプロフィールを利用することによって、ユーザに関係のない情報資源を削除することができ、その結果、フィルタリングの精度を向上させることができる。

【0069】《具体例2》具体例2は、ソーシャルフィルタリング（social filtering）あるいは協調的フィルタリング（collaborative filtering）と呼ばれるフィルタリングの機能を用いてユーザ間の相関関係を判定するようにしたものである。

【0070】ここで、ソーシャルフィルタリングとは次のようなフィルタリングである。即ち、情報フィルタリングの要求者であるユーザと、他の複数のユーザが、同一の情報資源に対して評価を行い、この評価値に基づき、他の複数のユーザのうち、要求者であるユーザと類似度の最も高いユーザを抽出し、情報フィルタリングの要求者であるユーザが未評価である情報資源に対して、類似度の最も高いユーザがこの情報資源をどう評価したかという情報を用いて、未評価の情報資源をフィルタリングするものである。

【0071】このソーシャルフィルタリングとは、ユーザの推薦情報や個人の好みの情報を集めて、それに基づいてフィルタリングする方法である。ソーシャルフィルタリングでは、映画の情報や音楽の情報のような、その

内容を計算機上で表現するのに難しいメディアに対して行われている。例えば、音、写真、芸術、ビデオ、マルチメディア情報などは、ユーザのプロフィールに関連した属性を見つけ、内容に基づくフィルタリングを行うのは難しい。そのため、そうした情報にはソーシャルフィルタリングが有効である。

【0072】具体的には、ソーシャルフィルタリングでは次のような処理を行う。

1. ユーザが、入力情報と情報資源（ここでは、本、ビデオ、音楽など）毎の評価情報を保存する。
2. 情報フィルタリング装置が、新しく登録されたユーザの評価結果と他のユーザの評価結果の相関関係を計算し、そのユーザがまだ見ていない情報資源（ここでは、本、ビデオ、音楽など）の評価値を予測する。
3. 新しい情報資源が登録されると、上記2.の相関関係を計算して、その情報資源に対するユーザの評価値を予測する。
4. 評価値の高い情報資源をユーザに提示する。
5. 新しい情報資源が登録される毎に、上記の2.、～4.の処理を繰り返す。

【0073】〈構成〉図7は、具体例2の構成図である。図の装置は、インタフェース管理部3、情報フィルタリング部4、メッセージ処理部5、情報提供部6、ユーザ評価情報格納部7、ソーシャルフィルタリング部8からなる。ここで、インタフェース管理部3～情報提供部6は、上記具体例1と同様である。

【0074】ユーザ評価情報格納部7は、任意のユーザがある情報資源に対して評価した情報を格納する。また、ソーシャルフィルタリング部8は、次のような機能を有している。

- 【0075】1. 情報資源毎にユーザの評価値をユーザ評価情報格納部7に保存する。
2. 上記1.で入力されたユーザの評価値と他のユーザの評価値の相関関係を計算し、そのユーザがまだ見ていない情報資源の評価値を予測する。
3. 評価値の高い情報資源をユーザに提示する。
4. 新しい情報資源が登録される毎に、上記の2.、3.を繰り返す。

【0076】このように、ソーシャルフィルタリング部8は、あるユーザのプロフィールに類似した、他のユーザのプロフィールを選ぶ機能を有している。更に、あるユーザの興味に一致しない他のユーザのプロフィールを見つける機能を有している。

【0077】〈動作〉具体例2においても、ユーザは、コンピュータの分野でデータベースの関係の雑誌の中で、「モデル」に関係する雑誌の記事をフィルタリングしたいとする。ユーザがこのフィルタリング装置にプロフィールを登録しておけば、情報フィルタリング装置は「モデル」に関係する雑誌の記事が、情報源10から情報提供部6に入ってくる毎に、その雑誌の記事をユーザ

に配送する。

【0078】図8は、具体例2の動作の主要部を示すフローチャートである。ここでは、具体例1におけるステップS10の後の状態であるとする。

【0079】〔ステップS1〕情報フィルタリング部4は、ユーザ20が興味がないといった記事に関して、その記事に興味があるユーザがいるかどうかというメッセージをメッセージ処理部5に送る。

【0080】〔ステップS2〕ソーシャルフィルタリング部8がステップS1のメッセージの返答をメッセージ処理部5に送り、適合する情報フィルタリング部4を探す。ここでは、ソーシャルフィルタリング部8が、情報フィルタリング部4 b~4 fを見つけたとする。図9は、その状態の説明図である。

【0081】〔ステップS3〕ソーシャルフィルタリング部8は、ユーザ20 aの興味に一致しない情報フィルタリング部4を見つける。これは、図9中に示す式

(7)によって、情報フィルタリング部4 b~4 fとの類似度を計算し、類似度が一番低いものがユーザ20 aの興味に一致しない情報フィルタリング部4を得る。ここでは、ユーザの興味に一致しない情報フィルタリング部4が、情報フィルタリング部4 cであったとする。

【0082】ここで、「ユーザ20 aの興味に一致しない情報フィルタリング部4の見つけ方」について説明する。まず、あるユーザの評点 U_i と他のあるユーザの評点 U_j の類似度を、例えば、図9中の式(7)から計算することができる。尚、この詳細については、文献[2]の中でも示されている。

【0083】図中の r_{ij} は類似度を示し、また、上線付きの U_i 、 U_j は、それぞれユーザの評点の平均、他のユーザの評点の平均を表している。ここで、評点とは、ユーザがある情報資源につけた興味の度合いを示す。この興味の度合いは、例えば、評点を、1,2,3,4,5,6,7という数字で付けられ、1がユーザがその情報資源に対して全く興味がないことを示し、7がユーザがその情報資源に対して非常に興味があることを示す。

【0084】このような式は、ポアソンrアルゴリズム(Pearson r Algorithm)と呼ばれている。上記式で用いた r_{ij} を全ての他のユーザに適用し、各ユーザ間の興味の近さを求める。

【0085】〔ステップS4〕ステップS3によって、見つけた情報フィルタリング部4を用いてフィルタリングを行う。ここでは、情報フィルタリング部4 aは、情報フィルタリング部4 cのプロファイルによってフィルタリングされる情報資源を除いてフィルタリングを行う。この時のフィルタリングの順番は次のようになる。

【0086】図10は、フィルタリングの順番を示す説明図である。図示のように、情報提供部6より提供された情報源10の文書識別子(情報資源)に対して、先ず、情報フィルタリング部4 cのプロファイルによるフ

ィルタリング(除外のフィルタリング)を行い、次に、情報フィルタリング部4 aのプロファイルによるフィルタリングを行い、そのフィルタリング結果を、インタフェース管理部3 aを介してユーザ20 aに出力する。

【0087】〈効果〉以上のように、具体例2によれば、ユーザのプロファイルの中の基底語で、その基底語を違う意味として使っているプロファイルを、ソーシャルフィルタリングの機能によって探し、そのプロファイルを利用することによって、ユーザに関係のない情報資源を削除することができる。その結果、プロファイルを修正するためのフィードバックの繰り返し回数が少ない場合でも、フィルタリングの精度をよくすることができるという効果がある。

【0088】《利用形態》本発明は、上記具体例に限定されるものではなく、以下の変形例についても適用が可能である。

【0089】・情報提供部6の代わりに、通常の情報検索システムに接続しても適用可能である。

【0090】・情報提供部6の代わりに、複数の情報源(例えば、新聞社、出版社、通信社など)からテキストやイメージなどの記事の配送を受けるシステムに接続しても適用可能である。

【0091】・具体例1、2では、情報フィルタリングの手法にベクトル空間モデルを用いたが、情報検索で用いられている確率モデルを用いてもよい。

【0092】・具体例1、2では、ユーザのプロファイルを修正する手法として、関連フィードバックを用いたが、機械学習の手法を用いてもよい。

【0093】・具体例1、2では、情報源10が一つであったが、複数の情報源であっても適用可能である。

【0094】・具体例2で用いたポアソンrアルゴリズムの代わりに、ユーザ間の相関関係を求める他のアルゴリズムを用いてもよい。

【図面の簡単な説明】

【図1】本発明の具体例1による情報フィルタリング方法のフローチャートである。

【図2】本発明の具体例1による情報フィルタリング装置の構成図である。

【図3】ベクトル空間モデルの演算式の説明図である。

【図4】本発明の具体例1における情報フィルタリング部の動作の説明図である。

【図5】ユーザの分類を示す説明図である。

【図6】本発明の具体例1におけるフィルタリングの順番を示す説明図である。

【図7】本発明の具体例2による情報フィルタリング装置の構成図である。

【図8】本発明の具体例2による情報フィルタリング方法の要部のフローチャートである。

【図9】本発明の具体例2における情報フィルタリング部の選択処理の説明図である。

【図10】本発明の具体例2におけるフィルタリングの順番を示す説明図である。

【符号の説明】

- 1 ユーザ情報格納部
- 2 個人情報管理部
- 3 インタフェース管理部

* 4 情報フィルタリング部

7 ユーザ評価情報格納部

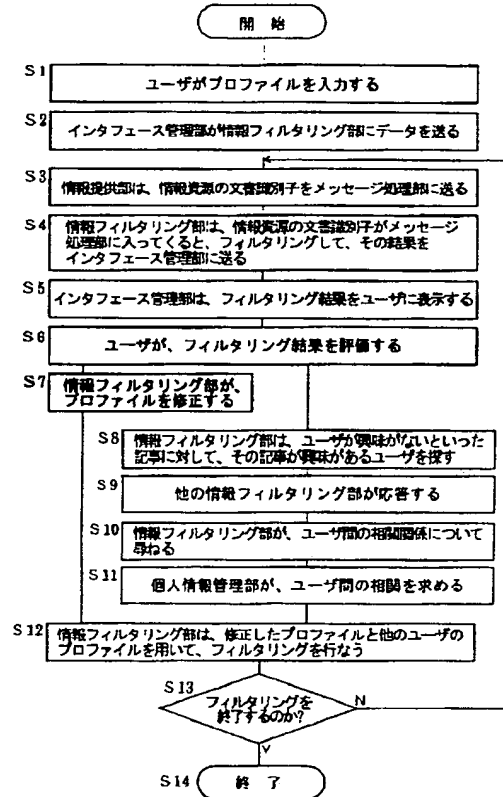
8 ソーシャルフィルタリング部

10 情報源

20 ユーザ

*

【図1】



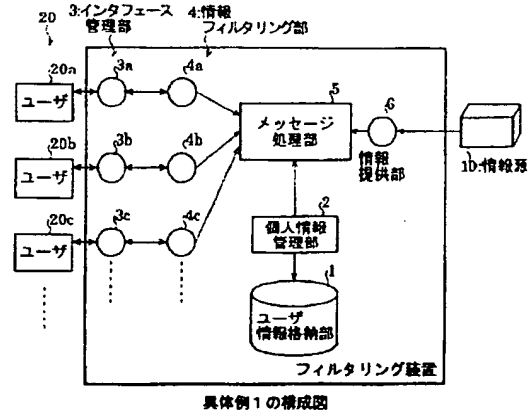
本発明方法の具体例1のフローチャート

【図3】

式 (1)	$DW = (dw_1, dw_2, \dots, dw_n)$
式 (2)	$Q = (q_1, q_2, \dots, q_n)$
式 (3)	$W_{ik} = \frac{(tf_{ik}) \cdot (\log \frac{N}{n_i})}{\sqrt{\sum_{j=1}^n (tf_{ij})^2 \cdot (\log \frac{N}{n_j})^2}}$
式 (4)	$\text{sim}(DW, Q) = \sum_{i=1}^n dw_i \cdot q_i$
式 (5)	$\text{sim}(DW, Q) > \theta$
式 (6)	$Q_i^{\text{new}} = \alpha Q_i^{\text{old}} + \beta \frac{1}{ \text{rel_docs}_i } \sum_{\text{rel_docs}_i} DW_i$ $-\gamma \frac{1}{ \text{nonrel_docs}_i } \sum_{\text{nonrel_docs}_i} DW_i$

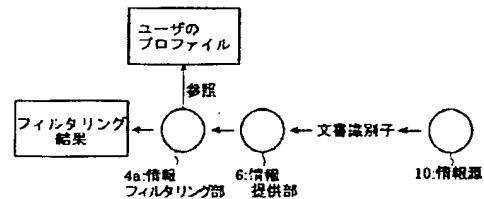
ベクトル空間モデルの演算式の説明図

【図2】



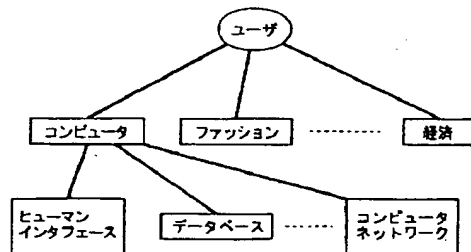
具体例1の構成図

【図4】



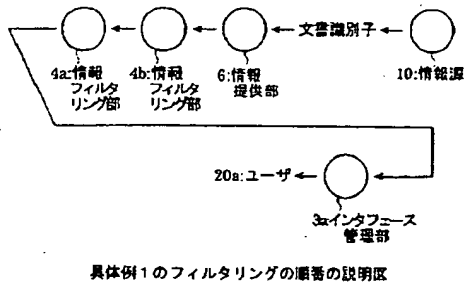
具体例1の情報フィルタリング部の動作の説明図

【図5】

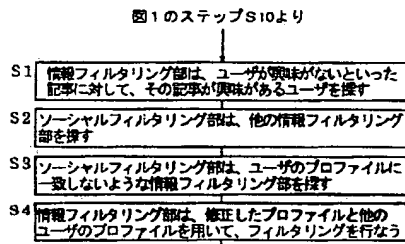


ユーザの分類を示す説明図

【図6】

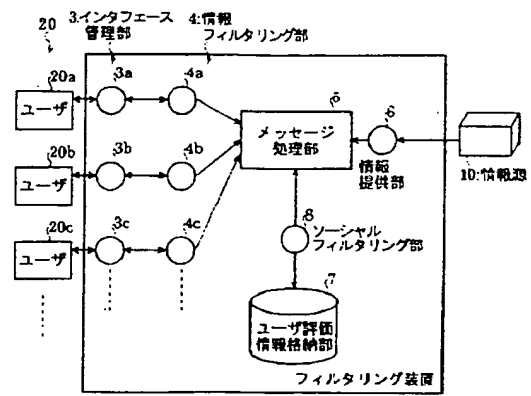


【図8】



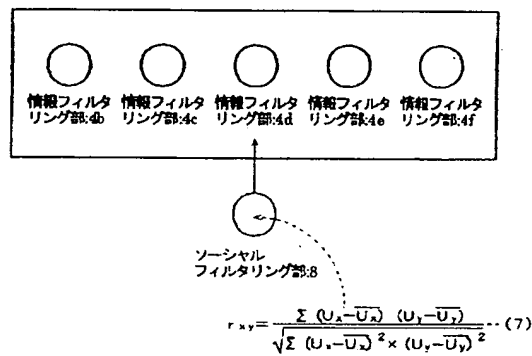
具体例2の要部のフローチャート

【図7】



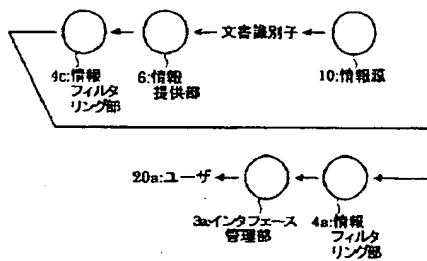
具体例2の構成図

【図9】



具体例2の情報フィルタリング部の選択処理の説明図

【図10】



具体例2のフィルタリングの順番の説明図